Supplemental Methods

Genetics of gene expression responses to temperature stress
in a sea urchin gene network

Contents

List of Figures

List of Tables

1 Gene expression analysis

1.1 qPCR

↩ To test for temperature and parental effects on genes known to be involved in environmental stress responses, we measured the expression of three genes: a control gene (*RBM8A*) and two chaperones (*Hsp70* and *Hsp90*) in the 12C and 18C treatments of the first experimental replicate with qPCR. These chaperone genes were not included on the DASL array (see below). qPCR primers are listed in Table S1. RNA samples were converted to cDNA using the High Performance cDNA Archive kit [Applied Biosystems]. We used the Qiagen SYBR-Fast kit on a Mastercycler ep realplex2 Thermal Cycler [Eppendorf] for all assays. Primers were tested for correlation ($R^2 > 0.99$) and efficiency ($95\% - 105\%$) using an 8-point standard curve of cDNA from a similarly staged larval culture. All qPCR measurements were run in triplicate with both chaperone genes and the control run on the same plate. Raw C_t scores are provided in Table S2. Samples were dropped if the standard deviation of C_t scores of either the focal gene or the reference gene was > 0.5. This left 44 and 42 samples, for *Hsp70* and *Hsp90*, respectively, with nearly equal numbers of samples at 12C and 18C and from each of the four male and four female parents. Relative expression values were calculated as:

$$R_i = \frac{E(\text{focal})^{\Delta C_t(\text{focal})}}{E(\text{ctl})^{\Delta C_t(\text{ctl})}}$$

for sample i, were E(focal) is the amplification efficiency of the focal gene, E(ctl) is the amplification efficiency of *RBM8A*, and the ΔC_t are $C_t(\text{ref}) - C_t(\text{focal})_i$ and $C_t(\text{ref}) - C_t(\text{ctl})_i$, respectively, with the $C_t(\text{ref})$ chosen as the mean C_t across the whole experiment. This calculation is based on the methods of Pfaffl (2001) and Hellemans et al. (2007) for normalization and minimizing variation, respectively. Tests for differential expression were performed by ANOVA on the $\log_2 R$ values. Models were selected from a full model with temperature, female and male parent effects and all pairwise interactions by backwards model selection, removing effects until all remaining effects were significant ($\alpha = 0.05$).

1.2 RNA-seq by SOLiD

1.2.1 Library preparation

↩ We chose seven samples representing three female and two male parents (Table S3) and used RNAseq on the SOLiD 3 plus platform (Life Technologies) to characterize transcriptome-wide differences in gene expression between embryos grown at 12C and 18C. $6.5 - 8.5\mu$g of DNase-treated total RNA was used as the input sample. mRNA was isolated via two rounds of poly-A selection using 100μl of Dynabeads (Invitrogen). Barcoded SOLiD libraries were prepared using the RNA-seq protocol for the SOLiD Total RNA-Seq kit and the SOLiD Transcriptome Multiplexing Kit (Applied Biosystems). Library quality was assessed using a 2100 Bioanalyzer (Agilent). Libraries were quantitated using qPCR, and prepared for standard stand-specific SOLiD sequencing in equimolar concentrations. 50bp single-end reads were generated on 2 slides by the IGSP Genome Sequencing and Analysis Core Resource at Duke University.

1.2.2 Read mapping

↩ Reads were pre-processed to separate barcoded samples and then mapped to the *S. purpuratus* genome v3.1 (Sodergren et al., 2006, `www.SpBase.org`) using bowtie (Langmead et al., 2009, v0.12.7) with the following parameters: (-C -n 2 -l 9 -e 100 -M 1 -t -best -trim3 15).

1.2.3 Gene model curation

↩ Reads were counted in gene models based on the gff3 file “GLEAN-3.1.gff3-chado-UTR” both downloaded from `www.SpBase.org`. We cleaned this gff3 file in two ways to improve the accuracy of our transcriptome estimates of gene expression differences: First, we converted all features of type “3UTR” to ”exon” so that reads mapping within annotated UTR regions would be counted towards that gene. Second, there are many cases in which one gene model is nearly, or entirely covered by a gene model with a different name. Reads mapping to the genome in locations covered by exons of multiple genes are discarded by *HTseq* (`http://www-huber.embl.de/users/anders/HTSeq`)since they cannot be unambiguously mapped to a single gene. These redundant gene models increased the apparent number of genes in the sea urchin genome, while reducing the number of genes that we could measure by RNAseq. First, we removed duplicate features with the same *seqid* using the UNIX commands *sort* and *uniq*. Next, we checked each gene model to see if it was overlapped by another gene model using the *coverageBed* program in the *BedTools* suit (Quinlan & Hall, 2010). If more than 70% of a gene model was overlapped by another gene model, we identified this second gene with the *BedTools* program *intersectBed* and kept only the larger of the two gene models. If multiple gene models overlapped each other under these criteria, we kept only the largest. If two gene models completely overlapped and had identical lengths, we chose one randomly. After this procedure, every gene model was unique over at least 30% of its length. This removed in total 2,857 of the original gene models.

1.2.4 Analysis

↩ We generated an average of 37 million reads per sample and counted the reads mapping uniquely to each of 26,428 gene models described above with *HTseq*. On average, 72% of the reads mapped to the genome (54% uniquely) and 31% mapped uniquely to one gene model. Since counting variance is much higher for low-expressed transcripts, we excluded all gene models with fewer than an average of 10 reads per sample. This left a total of 14,454 genes, which we then tested for an effect of temperature on gene expression. We assessed differential expression by culture temperature condition using the *R* (v1.13.1, Team 2011) package *edgeR* (Robinson et al., 2010, v2.2.6). Samples were normalized using the TMM method using default parameters. We estimated a common trended dispersion, and then tagwise dispersions for each gene using default parameters. For each of the 14,454 gene models, we fit a negative-binomial link generalized linear model with factors representing the male parent, female parent and temperature of the culture. We tested for an effect of temperature by comparing the full model containing all of these effects (male parent, A vs. D; female parent, A vs. C vs. D; and temperature: 12C vs. 18C) against the same model without the temperature factor (reduced model) using likelihood ratio tests. We controlled the false discovery rate (FDR) using the method of Benjamini & Hochberg (1995) implemented in the R function p.adjust. Genes were deemed significantly differentially expressed by temperature if their adjusted P-value was less than 0.05. Categorical enrichments of differentially expressed genes were performed based on the mappings of *S. purpuratus* genes to ENSEMBL proteins of Oliver et al. (2010). Tests for significant categorical enrichments were performed using the *wilcoxon.py* script of the python package *pyEnrichment* (`www.duke.edu/~ofedrigo/Olivier_Fedrigo/PythonScripts.html`), which tests if P-values for differential expression for genes within a given category are smaller than expected given p-values for all other genes. For comparisons among genes, *edgeR logConc* values were converted into Fragments Per Kilobase per Million mapped reads (RPKM) as RPKM $= \exp(\mu)/\text{length} \times 10^6$, where length is the sum of the exon lengths in kilobases and μ is the model intercept calculated by *edgeR*.

1.3 DASL Assay

↩ We used a DASL array (cDNA-mediated annealing, selection, extension and ligation, Illumina) to measure the transcript concentrations of 73 genes in each of 192 samples (Kuhn et al., 2004). The DASL platform is a method for measuring the expression of a select set of genes in a large number of samples (Kuhn et al., 2004). This platform is based on the technology underlying the GoldenGate genotyping platform produced by Illumina. Briefly, the platform uses the following steps: A custom assay pool is created by designing pairs of primers that target exons of the desired transcripts. The primer pairs both have universal primer extensions, and the downstream primer also contains a gene-specific capture sequence. When the primers hybridize to target cDNAs in solution, a PCR amplification using the universal primers joins a fluorescent-labeled primer to the gene-specific capture sequence. Individually labeled bead-types then capture fluorescent PCR products based on these capture-sequences, and then fluorescence associated with each bead is measured.

While sharing some features with any hybridization-fluorescence based expression technology, DASL differs from other, more common, microarray platforms in a number of ways that necessitate the use of different methods of background correction, summarization and normalization (Wong et al., 2008). For example, DASL uses many (about 30 on average) individual beads per sample for each probe set, rather that the typical 1-3 on other microarray platforms, and rather than using a mismatch probe-set, as is used in Affymetrix arrays, background normalization relies on a set of 27 bead-types with no complement in the target genome.

Like all platforms for measuring gene expression, there are quality control and normalization steps that must be taken before the gene expression data can be used in subsequent analyses. Our principle concern was to remove artifactual biases that might induce correlations among measures of different genes, or among measures of the same gene in different genetic backgrounds, as these biases would affect our downstream genetic analyses. To ensure a high standard of quality in the data, we wrote a customized pipeline in R (Team, 2010) for processing the raw data using many of the classes and methods from beadarray package of Bioconductor (Dunning et al., 2007; Gentleman et al., 2004). Our pipeline included five steps. 1) Identify and mask spatial artifacts within each array using the BASH functions of beadarrray (Cairns et al., 2008). 2) Correct for background variation among samples using the control probe distributions. 3) Identify poor quality samples 4) Test for consistency among probes targeting the same gene. 5) Normalize intensities among samples to control genes. All analysis was based only on the green channel intensities. Each of these steps is described in more detail below.

We worked with Illumina to design a custom DASL assay to measure the expression of the genes in the sea urchin embryogenic developmental gene regulatory network. We designed 384 probes to target annotated exons of 77 genes based on annotations to build 2.1 of the *S. purpuratus* genome on SpBase (`www.spbase.org`). Where possible, we validated the sequences from the sea urchin full genome sequence against targeted sequencing efforts available in GenBank. We chose 3-6 probes with Illumina Final scores > 0.8 (App version 6.4.1.0.0.0:2.0.0) for each gene. Illumina recommends using 3 probes per gene to improve the precision of each gene expression measurement. We included more probes when possible so that we could identify poorly performing probes based on the correlations of all probes targeting the same transcript. The set of probes that survived quality control steps (see below) and the sea urchin genes they target are listed in Table S4.

1.3.1 Mask problem areas in each array

↩ Array-based formats can suffer from spatial artifacts - regions of the assay surface that produce consistently different intensity readings. These may be due to camera or laser shadows, human error in loading the samples, or inherent biases around the edge of the array. If not accounted for, these effects can badly skew resulting analyses, even in cases like the DASL assay where each gene expression level is interrogated by a large number of probes (Cairns et al., 2008).

We used the adaptation of Harshlight to Illumina BeadArrays, BASH, implemented by (Cairns et al., 2008) in the Bioconductor package, beadarray (Dunning et al., 2007) with the following modified parameters: bgcorr=median as recommended for SAM arrays, diffsig = 0.001, and no imputation within outlier regions.

The BASH algorithm relies on the variance in expression measures within each bead-type as the statistic to identify spatial effects. We used log2-transformed values to identify outlier beads for this analysis, as recommended by (Cairns et al., 2008). We have found that log2-transformed intensities better identify outliers at the lower end of the distribution than does Iluminas recommended procedure. We chose an outlier cutoff of 2 mean absolute deviations (MADs) from the bead median for each bead-type in each sample. This is more conservative than the beadarray default and Illumina recommendation of 3 MADs. However, we find that this tends to eliminate more spurious beads.

Although some aspects of masks appear to be shared by most wells of the same plate (batch of 96 samples), most spatial artifacts appear to be individual well-specific. Thus we chose not to apply the same mask to all wells of a plate. We also observed that certain rows and columns within a plate suffered more from imaging artifacts than other wells. Overall, 8.4% (201072 of 2400798) of beads were removed by the masks. Of the remaining beads, about 20% were removed based on the 2 MAD from the median bead-specific cutoff. No probes ended up with less than 5 beads in any sample, and thus no probes were removed based on too few measures within a sample.

1.3.2 Background correction

↩ Background correction involves correcting for non-specific hybridization and for differences in the camera intensity among arrays. Differences in background levels among samples, if left uncorrected, lead to strong positive correlations between low expressed genes.

The DASL assay includes 27 non-specific bead-types intended to measure the background intensity of an array. We controlled for background variation and other technical artifacts using a principle-components regression strategy using the control beads. Since the intensity distribution of target and control probes was considerably different among the two 96 sample assay plates, we performed this background correction separately for each plate. For each plate, we performed a principle components analysis (PCA) on the intensity measurements of the 27 control probes across the 96 samples using the *prcomp* function in *R*. For the two plates, three and two axes, respectively each explained more than 80% of the total variation in control intensities. Under the assumption that biological differences among samples are unlikely to affect control probe intensities, patterns of target probe intensity that are correlated with these important PCA axes are likely due to technical artifacts in the DASL assay and can safely be removed. For each of the 384 target probes on each assay plate, we fit a linear model with the important PCA axes as predictors, and used the residuals from these models as "background corrected" data in downstream analyses.

1.3.3 Sample quality control

↩ To identify outlier samples, we measured the pairwise Pearson correlations among all pairs of samples. All but one sample had a correlation with at least one other sample greater than 0.90. The largest correlation between this particular sample and all other samples was 0.59. Thus, this sample was deemed sufficiently different from all other samples and we removed it from all further analyses.

1.3.4 Tests for probe consistency

↩ Each of the 384 probes was measured simultaneously in each sample, providing 3-6 independent estimates of the concentration of each target mRNA. Differences in intensity among probes targeting the same transcript may be due to inherent chemical differences among probes, splicing events between probes on the transcript, annotation errors of portions of these genes in the *S. purpuratus* genome, or simply stochastic noise. Fixed intensity differences among probes did not affect our analysis because all inference was based on deviations away from the overall probe mean. However, we flagged cases where the dynamics of different probes targeting the same transcript were distinct, suggesting splicing or annotation problems. To do this, we used data from a previous experiment using the same DASL array on 72 *S. purpuratus* cultures measured at seven stages throughout the embryonic period. For each of the 77 genes, after normalization and background correction, we measured the pairwise-correlations among the 3-6 target probes, and compared these values to their pairwise correlations with all other probes. Probes were kept if at least one of their correlation measures with the other probes on the target gene were among the top 5% of their correlations with all probes. Only genes with at least two remaining probes were kept in the analysis. This filtering step resulted in the removal of 44 probes, leaving 2-6 probes targeting 73 genes (Table S4), and improved our confidence that all remaining probes measured the intended transcripts.

1.3.5 Normalization

↩ We normalized samples by subtracting the average (log2) intensity of the four probes targeting the gene *RBM8A*. This gene was selected among three candidate normalization genes as part of an parallel study (Garfield et al, *in prep*) based on its consistency in expression across development, and the fact that it was not part of the focal gene regulatory network. If this gene were variable across samples, or variable in expression according to male or female parent, this normalization would induce correlations among all other genes. However, since it is outside of the network, these correlations are unlikely to be stronger among directly interacting genes. To explore this potential bias, we also tested two other independent normalization genes (*CyclinT* and *SoxB1*, the latter is part of the network but only at earlier developmental stages), and with all three genes jointly, and the reported patterns of male effect correlations between interacting genes were consistent in all three cases (Table S8).

The measured expression of *RBM8A* did decline at 18°C according to the RNAseq assay with a log2FC of -0.19, although this effect was not significant. Since the RNAseq assay targets many more genes, normalization factors calculated for these data are likely much more robust. Therefore, *RBM8A* may not be ideal for normalization across temperatures. To correct for this bias, we selected the four genes measured by DASL with the smallest temperature log2FC in the RNAseq assay (*Chordin*, *Not*, *Fmo2* and *FoxN2/3*), and subtracted the mean expression of all five genes across all cultures at each temperature from each sample. This adjustment made the estimated temperature effects across the two platforms much more similar (Fig. S2B). However, with or without this adjustment, we observed no correlation of temperature responses among directly interacting genes in the focal gene regulatory network.

1.3.6 DASL repeatability

↩ To assess the overall quality of the DASL assay, we quantified the Pearson correlations among technical and biological replicates. 16 technical replicates of the same sample were run at an earlier time. These values were all high: > 0.95 for technical replicates, > 0.87 for biological replicates.

2 Quantitative genetics of gene expression

↩ The observed variation among cultures in the expression of each of these genes reflects biological differences induced by the experimental design (temperature and male and female parent), by random differences in environments among individuals and cultures (spatial effects, embryo density, water condition, embryo stage and health), and measurement noise in the DASL system. Our experimental design allowed us to effectively isolate the effects of temperature and parentage from the random sources of error.

Among-individual variation was not assessed because not enough RNA could be collected per embryo. Instead, transcript levels were measured at the level of whole cultures, comprising hundreds of larvae, effectively averaging expression across full-sib cohorts. This feature of our experiment precludes the possibility of assessing narrow-sense heritability ($h^2 = \frac{V_A}{V_P}$) of expression since phenotypic variation among individuals (V_P) is unknown. However, it does not affect estimates of the additive genetic variance (V_A) or other measures of evolvability such as $I_A = \frac{V_A}{\bar{X}^2}$ (Houle, 1992).

2.1 Quantitative Genetic Model

2.1.1 Model Specification

↩ To quantify the effects of the environment (temperature treatment), genetic background, and other parental differences on the expression of each of these genes, we designed a Bayesian hierarchical mixed effect model. The design and interpretation of this model is based on the description of the North Carolina II breeding design in Lynch & Walsh (1998), while the prior structure and Gibbs algorithms are built on the treatments of Sorensen & Gianola (2010) and Hadfield (2010).

For each of the 73 genes, the observed data is $\mathbf{y}$, a $(r * 191) \times 1$ vector representing the probe intensities of the r (2-6, depending on gene) probes in each of the 191 samples that passed earlier quality control steps. Our interest is in overall transcript expression in each sample, which we treat as an unobserved latent variable, $\mathbf{u}$, and estimate its value based on the r probes. We model the effects of temperature, parents and developmental stage on these estimated sample expression values. Our hierarchical model can be represented as:

$$\begin{aligned} y_{i,j,m,n} &= \mu_i + u_{j,m,n} + e_{i,j,m,n} \\ u_{j,m,n} &= f(T_j, S_j, \mathbf{b}, M_m, F_n, D_{m,n}) + \epsilon_{j,m,n} \\ (\mu' \ \mathbf{b}' \ M'_m \ F'_n \ D_{m,n})' &\sim \pi(\theta) \end{aligned} \tag{1}$$

where $i = \{1, \ldots, r\}$ indexes a particular probe, $j = \{1, \ldots, 191\}$ a culture, $m = \{1, \ldots, 8\}$ a male parent and $n = \{1, \ldots, 8\}$ a female parent. Here, $\mu_i, i = \{2, \ldots, r\}$ is the fixed intensity difference for probe i across all cultures (μ_1 fixed at zero for identifiability of the $u_{j,m,n}$); T_j and S_j are the culture's temperature and mean developmental stage, both known variables; $\mathbf{b}$ are the fixed effects for temperature and stage; M_m and F_n are (2×1) vectors of the male and female "breeding values" and parent x temperature effects for parents m and n for the gene of interest; and $D_{m,n}$ is the interaction effect between male m and female n.

The function f specifies the linear mixed effect gene-by-environment model for gene expression:

$$f(T_j, S_j, \mathbf{b}, M_m, F_n, D_{m,n}) = \mathbf{b}' \begin{pmatrix} 1 \\ T_j \\ T_j^2 - (\overline{T_j^2}) \\ S_j \end{pmatrix} + M'_m \begin{pmatrix} 1 \\ T_j \end{pmatrix} + F'_n \begin{pmatrix} 1 \\ T_j \end{pmatrix} + D_{m,n} \quad (2)$$

Values of the covariates T_j were shifted by scale and location transformations to the set: $\{-0.5, 0, 0.5\}$, corresponding to the three experimental temperatures: 12°C, 15°C and 18°C. Thus, the parameter b_2 represents the change in expression from 12°C to 18°C, and $(b_1\ b_2\ b_3)$ models gene expression as a quadratic function of temperature. Priors on all model parameters, $\pi(\theta)$, are specified below.

The model can be written in matrix form as:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}_Y \begin{pmatrix} \mu \\ \mathbf{u} \end{pmatrix} + \mathbf{E} \qquad (3) \\ \mathbf{u} &= \mathbf{X}_U \mathbf{b} + \mathbf{Z}_M \mathbf{a} + \mathbf{Z}_F \mathbf{f} + \mathbf{Z}_D \mathbf{d} + \epsilon \end{aligned}$$

This notation roughly follows Schaeffer (2004). Here, $\mathbf{X}_Y$ is the design matrix relating fixed probe effects (μ) and latent transcript expression ($\mathbf{u}$) to observed probe intensities. $\mathbf{X}_U$ is the fixed effect design matrix for the temperature and developmental stage effects, and $\mathbf{Z}_M$, $\mathbf{Z}_F$ and $\mathbf{Z}_D$ are the random effect incidence matrices. $\mathbf{a} = (M'_1\ M'_2\ \dots\ M'_8)'$ is a vector of male effects on expression (breeding values). The two elements for each male are the intercept and slope of a linear function of temperature - the random regression coefficients. $\mathbf{f} = (F'_1\ F'_2\ \dots\ F'_8)'$ is similarly structured for female effects, and $\mathbf{d}$ is the vector of parent interactions. The matrices $\mathbf{Z}_M$ and $\mathbf{Z}_F$ are modified incidence matrices which relate the random regression of male or female effects on temperature to $\mathbf{u}$. Each consecutive pair of columns has rows: $(0 \dots 0\ 1\ T_j\ 0 \dots 0)$ which is non-zero only in the elements corresponding the the male (female) parent of the sample.

The fixed effects, μ and $\mathbf{b}$ are assigned diffuse independent normal priors:

$$\begin{pmatrix} \mu \\ \mathbf{u} \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, 10^3 \times \mathbf{I}_{3+r-1}\right) \qquad (4)$$

where $\mathbf{0}$ is the zero vector and $\mathbf{I}_k$ is the k-dimensional identity matrix.

The random effects, **a**, **f** and **d** are assigned independent multivariate normal priors:

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{f} \\ \mathbf{d} \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{I}_8 \otimes \mathbf{G} & 0 & 0 \\ 0 & \mathbf{I}_8 \otimes \Sigma_F & 0 \\ 0 & 0 & \sigma_d^2 \times \mathbf{I}_{32} \end{bmatrix}\right) \tag{5}$$

where $\mathbf{G}$ and Σ_F are 2×2 covariance matrixes for the male and female effects, respectively, σ_d^2 is the variance of interaction effects, and the symbol $\otimes$ represents the Kronecker product. The covariances $\mathbf{G}$ and Σ_F allow non-zero covariance between the intercept and slope coefficients of the random regression functions of temperature for the male and female effects.

The noise terms, $e_{i,j,m,n}$ and $\epsilon_{j,m,n}$ are assigned t-distributions (given σ_Y^2 and σ_U^2, respectively) with the following hierarchical specification:

$$\begin{aligned} e_{i,j,m,n} &\sim \mathrm{N}(0, \sigma_Y^2/\lambda_{i,j,m,n}) \\ \epsilon_{j,m,n} &\sim \mathrm{N}(0, \sigma_U^2/\delta_{j,m,n}) \\ \lambda_{i,j,m,n} &\sim \mathrm{Ga}(a_Y, b_Y) \\ \delta_{i,j,m,n} &\sim \mathrm{Ga}(a_U, b_U) \\ 1/\sigma_Y^2 &\sim \mathrm{Ga}(g_Y, h_Y) \\ 1/\sigma_U^2 &\sim \mathrm{Ga}(g_U, h_U) \end{aligned} \tag{6}$$

with $\mathrm{Ga}(a, b)$ the gamma distribution with shape a and rate b. This prior specification allows heavier tails than the commonly used normal distribution for the model residuals. This reduces the influence of “outlier” measurements, for example where particular probe measurements are far from expected given other probes in the same sample. We found this distribution improved the stability of the estimates of transcript levels and parent effects, given the often noisy DASL measurements.

We placed inverse Wishart priors on the two covariance matrices of random effects, and a gamma prior on the inverse random interaction effect variance:

$$\begin{aligned} \mathbf{G} &\sim \mathrm{iW}(S, v) \\ \Sigma_F &\sim \mathrm{iW}(S, v) \\ \sigma_d^2 &\sim \mathrm{Ga}(a_d, b_d) \end{aligned} \tag{7}$$

with $\mathrm{iW}(S, v)$ the inverse Wishart distribution with v degrees of freedom, and inverse scale matrix S.

2.1.2 Implementation

We implemented a Gibbs sampler in R v1.13.1 to fit this hierarchical mixed effects model. In sequence, we updated the random and fixed effects and variances of the probe and sample level models. The posterior of all parameters is:

$$\begin{aligned} p(\mu, \mathbf{u}, \mathbf{b}, \mathbf{a}, \mathbf{f}, \mathbf{d}, \mathbf{G}, \Sigma_F, \sigma_d^2, \lambda, \delta, \sigma_Y^2, \sigma_U^2 \mid \mathbf{y}) \propto& \\ & p(\mathbf{y} \mid \mu, \mathbf{u}, \lambda, \sigma_Y^2) \\ & \times p(\mathbf{u} \mid \mathbf{b}, \mathbf{a}, \mathbf{f}, \mathbf{d}, \delta, \sigma_U^2) p(\mathbf{a}, \mathbf{f}, \mathbf{d} \mid \mathbf{G}, \Sigma_F, \sigma_d^2) \\ & \times \pi(\mathbf{b}, \sigma_Y^2, \sigma_U^2, \lambda, \delta, \mathbf{G}, \Sigma_F, \sigma_d^2) \end{aligned} \tag{8}$$

The update for the latent transcript expression variables, $\mathbf{u}$ is a draw from the multivariate normal density given by:

$$\mathbf{u} \mid \theta_{-u} \sim \mathrm{N}\left(\hat{\mathbf{u}}, \mathbf{C}^{-1}\right) \tag{9}$$

where θ_{-u} represents all model parameters except u and:

$$\begin{aligned} \mathbf{C} &= \mathbf{X}_Y' \Sigma_Y^{-1} \mathbf{X}_Y + \Sigma_U^{-1} \\ \hat{\mathbf{u}} &= \mathbf{C}^{-1} \left(\mathbf{y}' \Sigma_Y^{-1} \mathbf{X}_Y + (\mathbf{X}_U \mathbf{b} + \mathbf{Z}_M \mathbf{a} + \mathbf{Z}_F \mathbf{f} + \mathbf{Z}_D \mathbf{d})' \, \Sigma_U^{-1} \right) \\ \Sigma_Y &= \mathrm{diag}(\sigma_Y^2 / \lambda_{i,j,m,n}) \\ \Sigma_U &= \mathrm{diag}(\sigma_U^2 / \delta_{j,m,n}) \end{aligned} \tag{10}$$

and $\mathrm{diag}(a_i)$ is the diagonal matrix with entries a_i.

The parameters $\mathbf{b}, \mathbf{a}, \mathbf{f}$ and $\mathbf{d}$ are updated in a single block update by drawing from the multivariate normal density:

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{a} \\ \mathbf{f} \\ \mathbf{d} \end{pmatrix} \mid \theta_{-\mathbf{b},\mathbf{a},\mathbf{f},\mathbf{d}} \sim \mathrm{N}\left(\mathbf{r}, \mathbf{C}^{-1}\right) \tag{11}$$

where:

$$\begin{aligned} \mathbf{C} &= \mathbf{W}' \Sigma_U^{-1} \mathbf{W} + \Sigma_r^{-1} \\ \mathbf{r} &= \mathbf{C}^{-1} \left(\mathbf{u}' \Sigma_U^{-1} \mathbf{W} \right) \\ \mathbf{W} &= [\mathbf{X}_U \; \mathbf{Z}_M \; \mathbf{Z}_F \; \mathbf{Z}_d] \\ \Sigma_U &= \mathrm{diag}(\sigma_U^2 / \delta_{j,m,n}) \\ \Sigma_r &= (10^3 \times \mathbf{I}_3) \oplus (\mathbf{I}_8 \otimes \mathbf{G}) \oplus (\mathbf{I}_8 \otimes \Sigma_F) \oplus (\sigma_d^2 \times \mathbf{I}_{32}) \end{aligned} \tag{12}$$

$\oplus$ is the matrix direct sum, and $\otimes$ is the Kronecker product, as before.

Updates of the random effect (co)variance parameters are given as draws from inverse Wishart and inverse Gamma distributions:

$$\begin{aligned}
\mathbf{G} \mid \theta_{-\mathbf{G}} &\sim \mathrm{iW}\left((S + vec^{-1}(\mathbf{a})vec^{-1}(\mathbf{a})', (v+8)\right) \\
\Sigma_F \mid \theta_{-\Sigma_F} &\sim \mathrm{iW}\left((S + vec^{-1}(\mathbf{f})vec^{-1}(\mathbf{f})', (v+8)\right) \\
\sigma_d^2 \mid \theta_{-\sigma_d^2} &\sim \mathrm{iG}\left((a_d + 32/2), b_d + \frac{\sum_k^{32} d_k^2}{2}\right)
\end{aligned} \tag{13}$$

where iG(a, b) is the inverse Gamma distribution with shape a and rate b, and $vec{-}1$ is the inverse vectorization operator that takes (for example) the vector $\mathbf{a}$ and folds it into a 2×8 matrix. The numbers 8 and 32 in these equations refer to the number of male or female parents (8), and length of the vector $\mathbf{d}$, respectively.

Finally, updates for the residual parameters of the models for $\mathbf{y}$ and $\mathbf{u}$ consist of a series of independent draws from univariate gamma distributions:

$$\begin{aligned}
1/\sigma_Y^2 \mid \theta_{-\sigma_Y^2} &\sim \mathrm{Ga}\left(\left(g_Y + \frac{191 * r}{2}\right), h_Y + \frac{||\mathrm{diag}(\lambda_{i,j,m,n}^{1/2})\hat{\mathbf{r}}_Y||^2}{2}\right) \\
1/\sigma_U^2 \mid \theta_{-\sigma_U^2} &\sim \mathrm{Ga}\left(\left(g_U + \frac{191}{2}\right), h_U + \frac{||\mathrm{diag}(\delta_{j,m,n}^{1/2})\hat{\mathbf{r}}_U||^2}{2}\right) \\
\lambda \mid \theta_{-\lambda} &\sim \mathrm{Ga}\left(\left(a_Y + \frac{1}{2}\right), b_Y + \frac{(\hat{\mathbf{r}}_Y)^2}{2\sigma_Y^2}\right) \\
\delta \mid \theta_{-\delta} &\sim \mathrm{Ga}\left(\left(a_U + \frac{1}{2}\right), b_U + \frac{(\hat{\mathbf{r}}_U)^2}{2\sigma_U^2}\right)
\end{aligned} \tag{14}$$

with $||\mathbf{x}||^2$ the squared Frobenious norm, $(\mathbf{x})^2$ the element-wise square of the vector $\mathbf{x}$, and Ga$(a, \mathbf{b})$ a vector of independent gamma distributions all with scale a, but each with a different shape given by the appropriate entry of the vector $\mathbf{b}$. Here, the residual vectors are:

$$\begin{aligned}
\hat{\mathbf{r}}_Y &= \mathbf{y} - \mathbf{X}_Y \begin{pmatrix} \mu \\ \mathbf{u} \end{pmatrix} \\
\hat{\mathbf{r}}_U &= \mathbf{u} - (\mathbf{X}_U\mathbf{b} + \mathbf{Z}_M\mathbf{a} + \mathbf{Z}_F\mathbf{f} + \mathbf{Z}_D\mathbf{d})
\end{aligned} \tag{15}$$

2.1.3 Prior hyperparameters

↩ Prior hyper-parameters used in this paper are listed in Table S8. Scale values for the gamma priors of variance components were chosen so that these distributions were proper, but relatively uninformative. Since we had no prior expectation that genetic, environmental or residual variances should be higher for any particular gene, prior means for the variances were chosen proportional to the observed total variance of each gene. While not fully Bayesian since the prior relies on observed data, this procedure uses the data minimally and prevents the prior from having more influence on some genes than others. In particular, the prior mean for the variance of residuals of probe intensity (σ^2_Y) was chosen proportionally to the observed total variance of probe intensities, while the prior means for the variance of residuals of transcript expression (σ^2_U) and the diagonal elements of $\mathbf{G}$, Σ_F and σ^2_d were all chosen proportionally to the observed variance of mean probe intensities per sample. Prior means for the elements of λ and δ were set to one.

We tested the sensitivity of this analysis to prior choice by re-running all models and downstream analyses with several alternative choices of prior hyper-parameters. These alternative choices are listed in Table S8. In particular, we explored the sensitivity of model fits to each class of parameter, either by making the prior shapes more peaked over their mean, or shifting the prior means. Increasing the scale parameters of δ had no effect, while increasing the scale parameters of λ increased the appearance of male-by-female effects. This is consistent with the presence of strong outlier points biasing estimates of parameters with less data. The prior mean values for the male and female effect variances affected the magnitude of estimated male and female effects on expression, but did not change the relative ordering of male or female effects within or among genes. Setting the prior means to 10% of observed among-culture variance resulted in posterior means in a similar range, suggesting that the priors did not induce a strong, consistent bias among genes. The prior mean for the "fixed" effect variances had no effect for values of the exponent greater than ~ 2. Below this, the prior variance of these effects shrunk the fitted values towards zero.

2.1.4 MCMC estimation

↩ We performed posterior simulation by running a single MCMC chain for each gene, allowing a burn-in period of 10,000 iterations and then drawing 1,000 posterior samples of all parameters with a thinning rate of 10. We assessed convergence by measuring the autocorrelation of each parameter, and by re-running each chain from different starting values. All parameters appeared well converged by these measures. Posterior distributions were summarized as a mean and a credible interval spanning the central 95% of posterior samples. Male, female, temperature, parent-interaction and parent-by-temperature effects were tested for "significance" by inspecting the credible intervals for all parameters of a class (ex. breeding values for each of the eight male parents). If any individual effect, or temperature coefficient, had a credible interval that did not cross zero, we counted the class as being important. This is a relatively conservative test as it relies on at least one individual having a very large (or small) breeding value. For downstream analyses, we used posterior means as estimates of each effect.

References

Benjamini, Y. & Hochberg, Y. 1995. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *Journal Of The Royal Statistical Society Series B-Methodological* **57**: 289–300.

Cairns, J., Dunning, M., Ritchie, M., Russell, R. & Lynch, A. 2008. BASH: a tool for managing BeadArray spatial artefacts. *Bioinformatics* **24**: 2921–2922.

Dunning, M., Smith, M., Ritchie, M. & Tavare, S. 2007. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **23**: 2183–2184.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y.H. & Zhang, J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**: R80.

Hadfield, J. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software* .

Hammond, L.M. & Hofmann, G.E. 2010. Thermal tolerance of Strongylocentrotus purpuratus early life history stages: mortality, stress-induced gene expression and biogeographic patterns. *Marine Biology* **157**: 2677–2687.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**: R19.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195–204.

Kampstra, P. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software, Code Snippets* pp. 1–9.

Kuhn, K., Baker, S., Chudin, E., Lieu, M.H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T. & Chee, M. 2004. A novel, high-performance random array platform for quantitative gene expression profiling. *Genome research* **14**: 2347–2356.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.

Longabaugh, W.J.R., Davidson, E.H. & Bolouri, H. 2005. Computational representation of developmental genetic regulatory networks. *Developmental biology* **283**: 1–16.

Lynch, M. & Walsh, B. 1998. *Genetics and analysis of quantitative traits*, 1st edn. Sinauer Associates Inc.

Oliver, T.A., Garfield, D.A., Manier, M.K., Haygood, R., Wray, G.A. & Palumbi, S.R. 2010. Whole-genome positive selection and habitat-driven evolution in a shallow and a deep-sea urchin. *Genome biology and evolution* **2**: 800–814.

Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**: e45.

Quinlan, A.R. & Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Schaeffer, L. 2004. Application of random regression models in animal breeding. *Livestock Production Science* **86**: 35–45.

Sodergren, E., Weinstock, G., Davidson, E., Cameron, R., Gibbs, R., Angerer, R., Angerer, L., Arnone, M., Burgess, D., Burke, R., Coffman, J., Dean, M., Elphick, M., Ettensohn, C., Foltz, K., Hamdoun, A., Hynes, R., Klein, W., Marzluff, W., McClay, D., Morris, R., Mushegian, A., Rast, J., Smith, L., Thorndyke, M., Vacquier, V., Wessel, G., Wray, G., Zhang, L., Elsik, C., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M., Mackey, A., Maglott, D., Panopoulou, G., Poustka, A., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C., Worley, K., Durbin, K., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M., Jackson, A., Milosavljevic, A., Tong, M., Killian, C., Livingston, B., Wilt, F., Adams, N., Belle, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Geneviere, A.M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A., Goldstone, J., Cole, B., Epel, D., Gold, B., Hahn, M., Howard-Ashby, M., Scally, M., Stegeman, J., Allgood, E., Cool, J., Judkins, K., McCafferty, S., Musante, A., Obar, R., Rawson, A., Rossetti, B., Gibbons, I., Hoffman, M., Leone, A., Istrail, S., Materna, S., Samanta, M., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K.F., Brandhorst, B., Whittle, J., Berney, K., Bottjer, D., Calestani, C., Peterson, K., Chow, E., Yuan, Q., Elhaik, E., Graur, D., Reese, J., Bosdet, I., Heesun, S., Marra, M., Schein, J., Anderson, M., Brockton, V., Buckley, K., Cohen, A., Fugmann, S., Hibino, T., Loza-Coll, M., Majeske, A., Messier, C., Nair, S., Pancer, Z., Terwilliger, D., Agca, C., Arboleda, E., Chen, N., Churcher, A., Hallbook, F., Humphrey, G., Idris, M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R., Raible, F., Rowe, M., Taylor, J., Tessmar-Raible, K., Wang, D., Wilson, K., Yaguchi, S., Gaasterland, T., Galindo, B., Gunaratne, H., Juliano, C., Kinukawa, M., Moy, G., Neill, A., Nomura, M., Raisch, M., Reade, A., Roux, M., Song, J., Su, Y.H., Townley, I., Voronina, E., Wong, J., Amore, G., Branno, M., Brown, E., Cavalieri, V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Rottinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, G., Miranda, E., Thomason, R., Walton, K., Wikramanayke, A., Wu, S.Y., Xu, R., Brown, C., Chen, L., Gray, R., Lee, P., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S., Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L., Okwuonu, G., Parker, D., Pu, L.L., Thorn, R. & Wright, R. 2006. The Genome of the Sea Urchin Strongylocentrotus purpuratus. *Science (New York, NY)* **314**: 941–952.

Sorensen, D. & Gianola, D. 2010. *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics (Statistics for Biology and Health).* Springer.

Team, R.D.C. 2010. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing.

Wong, W., Loh, M. & Eisenhaber, F. 2008. On the necessity of different statistical treatment for Illumina BeadChip and Affymetrix GeneChip data and its significance for biological interpretation. *Biology Direct* **3**: 23.

Table S1. Primer sequences, amplification efficiencies and original citations for the qPCR assays used to assay *Hsp70* and *Hsp90* expression.↩

Gene	Primers	R^2	Efficiency	Citation
RBM8A	ATGAAGCCGAAGAGGATGAA	0.993	98.35	This study
	GACCCTGGAACTCCTCATCA			
Hsp70	AAGATATGAGGTCCAACCCAAGAT	0.999	96.2	[1]
	TGCTGAAGCACTGCTTGACA			
Hsp90	AGGAGGAAGCGATCAAACTG	0.982	108.82	This study
	TCATCATACGGTTGACCTCAG			

[1] (Hammond & Hofmann, 2010)

Table S2. Raw qPCR data on Hsp70 and Hsp90. Data file containing non-normalized CT scores for the two assayed chaperone genes, *Hsp70* and *Hsp90*, and the control gene, *RBM8A*. Assays were run in triplicate, and samples in which either the target gene or the control gene had a standard deviation of C_t scores > 0.5 were discarded. Control genes were only used if they were run on the same plate as the target genes. **File:** `qPCR_Ct_scores.dat`↩

Table S3. Sample info for RNA-seq. The eight samples chosen for RNA-seq analysis on the SOLiD 3 plus platform are shown. For each sample, the male and female parents and temperature treatment are listed, as well as statistics from the RNA-seq output. *Total reads* is the total count of reads reported from the run. *Mapped reads* is the number of reads that mapped somewhere in the *S. purpuratus* v3.1 genome. *Unique reads* had a unique best match to the genome according to Bowtie. *Gene reads* were reads that mapped uniquely to an exon of one gene model.↩

Sample	Female	Male	Temp	Total reads	mapped reads	unique reads	Gene reads
AA2-18	A	A	18C	32607339	22675408	16947450	9582895
AD1-18	A	D	18C	44858154	32380417	24068050	13556927
DA2-18	D	A	18C	40503392	30233196	22623508	15243065
CD1-18	C	D	18C	35453186	25666069	19105707	10818142
AD2-12	A	D	12C	32653859	24409674	18079737	10368740
CA2-12	C	A	12C	43882450	30361002	22447485	12543146
CD1-12	C	D	12C	32525751	24125691	17557885	10532719

Table S4. Annotation information for developmental genes assayed by DASL. Each of the 73 developmental genes plus controls analyzed are listed, along with official *S. purpuratus* gene IDs (SPU), a classification of gene function, whether the gene is involved in an active regulatory event during gastrulation and the embryonic territories in which the gene is expressed during gastrulation. Annotation information on gene activity was extracted from the 27-30h time points of the BioTapestry representations of the endomesodermal network (Longabaugh et al., 2005, `www.biotapestry.org`) and the ectodermal network from the Davidson lab website: `http://www.its.caltech.edu/~mirsky/`) as of September 05, 2011. **File:** `Network_gene_annotation.dat`↩

Table S5. Gene network relationships among target developmental genes. The 93 known regulatory events among 52 of the 72 assayed developmental genes are listed. For each regulatory event, the upstream and downstream gene are listed, as well as the timing of the event (hpf at 15C), its location within an embryo, and whether the downstream gene is promoted or repressed. Annotation information was extracted from the BioTapestry representations of the endomesodermal and ectodermal network as above. **File:** `Network_interactions_list.dat`↩

Table S6. Female parent effects on heat shock gene expression at 18C but not 12C. Both Hsp70 and Hsp90 had significant female parent effects at 18C, but not at 12C, and no significant male parent effects at either temperature. ANOVA tables for the model: log2(exp)ij = Femalei + Malej + eij are shown for each gene x temperature combination.↩

Hsp70 12C	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Female	3	8.88	2.96	0.52	0.6770
Male	3	4.46	1.49	0.26	0.8535
Residuals	15	85.95	5.73		

Hsp70 18C	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Female	3	20.24	6.75	6.25	0.0058
Male	3	8.00	2.67	2.47	0.1017
Residuals	15	16.20	1.08		

Hsp90 12C	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Female	3	2.50	0.83	0.47	0.7112
Male	3	2.93	0.98	0.54	0.6596
Residuals	14	25.12	1.79		

Hsp90 18C	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Female	3	11.71	3.90	6.82	0.0046
Male	3	2.03	0.68	1.18	0.3511
Residuals	14	8.01	0.57		

Table S7. Embryo morphology data. Embryo lengths and stages measured on 12-23 embryos per cultures. Culture info (Experimental Block, Female parent, Male parent, Temperature) is provided. **File:** `morphology_data.dat`↩

Table S8. Prior hyperparameters used in this study. $\sigma^2_{\hat{s}}$ is the variance of culture means, and $\sigma^2_{\hat{p}}$ is the residual probe variance (over all probes) after removing the culture means. Each alternative listed was tested individually by re-fitting the model for each gene, and re-testing the major results.↩

	prior values	alternatives tested	sensitivity of results
a_Y	5	100	More interaction effects significant
b_Y	a_Y		
g_Y	2		
h_Y	$\sigma^2_{\hat{p}} \times g_Y$		
a_U	100	5	none
b_U	a_U		
g_U	2		
h_U	$\sigma^2_{\hat{s}} \times g_U$		
S	$\frac{\sigma^2_{\hat{s}}(\nu-3)}{10}\mathbf{I}_2$	$\frac{\sigma^2_{\hat{s}}(\nu-3)}{3}\mathbf{I}_2$, $\frac{\sigma^2_{\hat{s}}(\nu-3)}{30}\mathbf{I}_2$	Magnitude of $\mathbf{G}$ and Σ_F increases, or decreases, respectively, but genetic variance still tends to be larger than temperature effects
ν	4	3.25	
a_d	2	5	
b_d	$\frac{10}{\sigma^2_{\hat{s}}(a_d-1)}$	$\frac{3}{\sigma^2_{\hat{s}}(a_d-1)}$, $\frac{30}{\sigma^2_{\hat{s}}(a_d-1)}$	Magnitude of male x female effects, changes, but not significance
σ^2_{fixed}	3	$1, 6$	none
normGene	*RBM8A*	*CyclinT*, *SoxB1*, (*RBM8A*, *CyclinT*, and *SoxB1*)	None for CyclinT. Male effect correlations were reduced with SoxB1 or all three genes together, but correlations in network were still significant

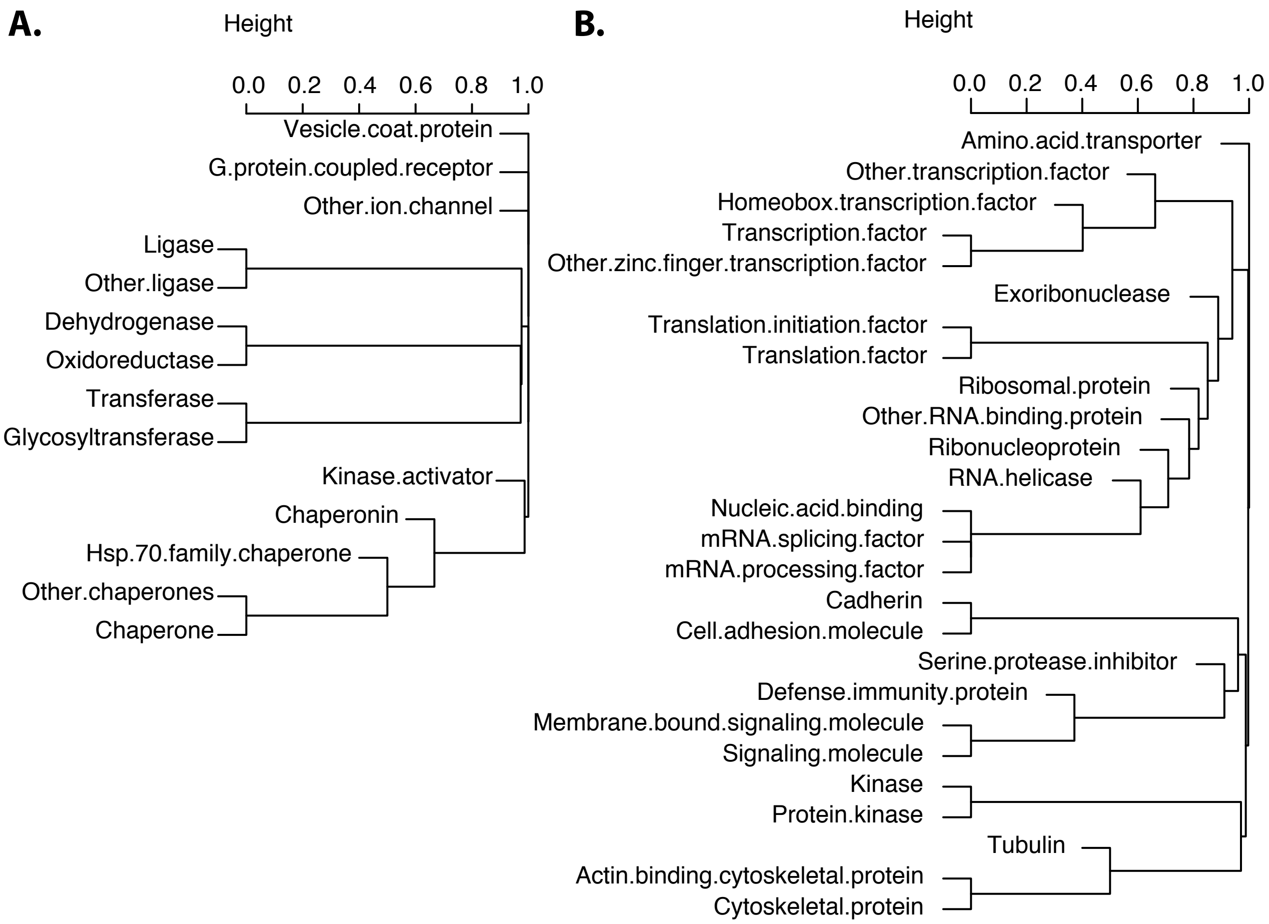

Figure S1. Dendrograms showing relationships among significantly enriched PANTHER categories for up- (A) and down- (B) regulated genes. Distances between two categories are calculated based on the proportion of the genes (with adjusted temperature effect $P < 0.05$) in the smaller category that are also part of the larger category. When two categories have a distance of zero, the smaller category is entirely contained within the larger category. All categories displayed were significantly enriched for up- or down-regulated genes with an adjusted enrichment $P < 0.05$.↩

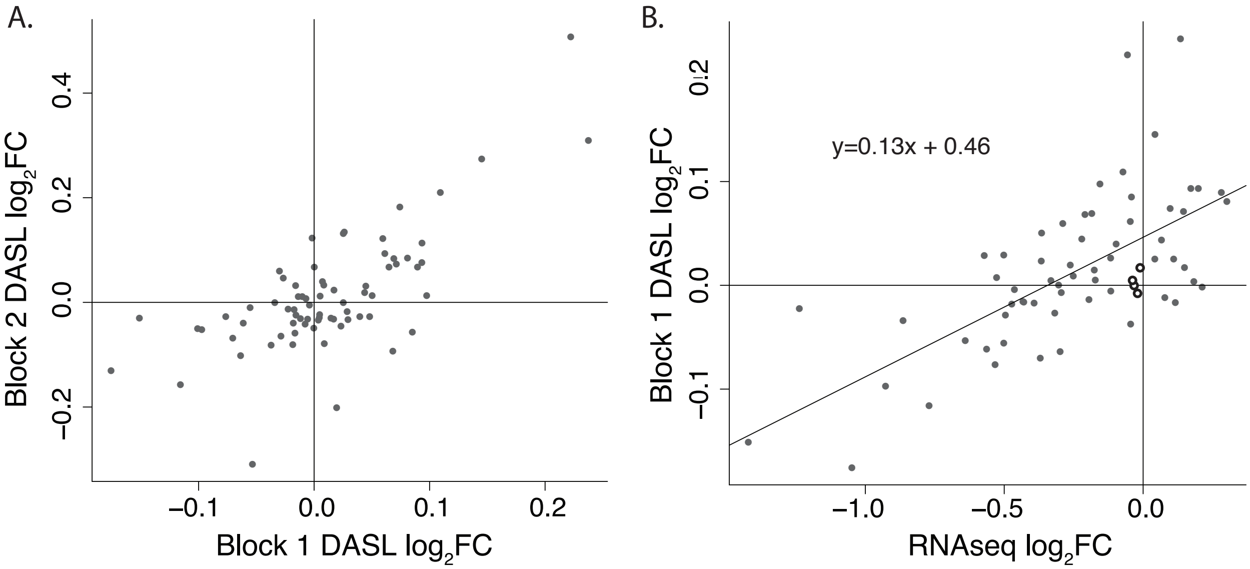

Figure S2. Estimates of temperature responses by RNA-seq and DASL were consistent. A) Estimated temperature responses across the two replicates of the experiment, as measured by DASL (r=0.71). **B)** Estimated log2 fold change (log2FC) between 12C and 18C for genes measured by both RNA-seq and DASL. The two estimates are well correlated (r=0.68) for DASL estimates from the first replicate of the experiment (all RNA-seq samples were from this run), but RNA-seq consistently estimated larger expression responses. The four genes used to re-normalize DASL expression across temperatures are shown as hollow circles. These genes were chosen as the five genes with the smallest responses according to RNAseq. The diagonal line is a least-squares regression of the DASL estimates on the RNA-seq estimates.↩

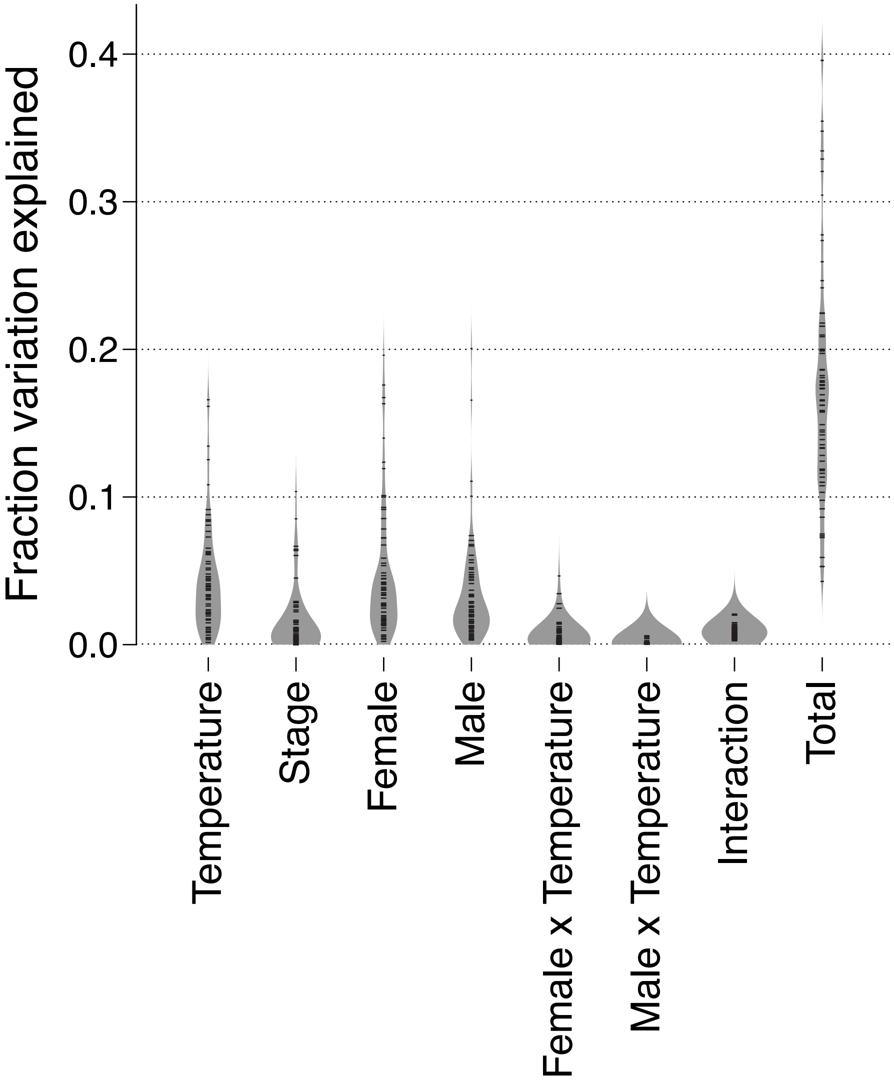

Figure S3. Percentage of variation in gene expression explained by temperature, stage and parental effects. Bean plots (Kampstra, 2008) showing the percent of the total observed variation accounted for by each of the modeled factors over the 72 genes. Greg curves show kernel densities and black dashes show the values for each gene. Total variance was calculated as the variance in mean intensities for each sample across the 2-6 probes that targeted each gene. The full model is described above.↩

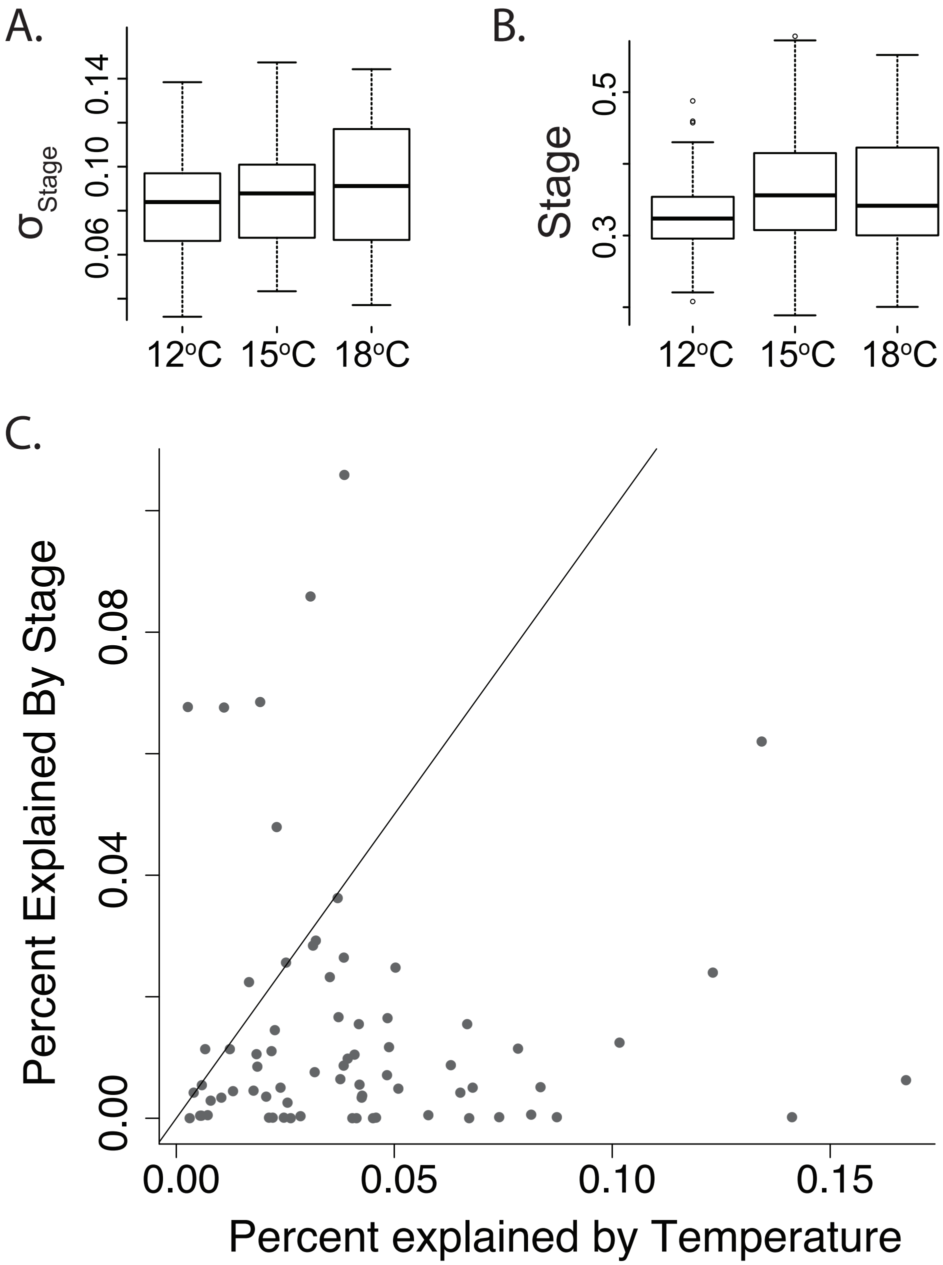

Figure S4. Cultures were slightly more variable at higher temperatures, and differences in sampling time by temperature was small. A) 18C and 12C temperatures induced a slight increase in the variability in developmental rate among embryos within each culture. Boxplots show median and quantiles of the distribution of the standard deviation of embryonic stages (proportion of the blastocoel traversed by the archenteron, see Fig. 1A) within cultures at each temperature. **B)** Mean stage at sampling was slightly higher at 15C (and less so at 18C) than 12C. **C)** Mean stage differences among cultures accounted for less gene expression variation than the temperature treatment for the majority of genes, including all genes with a significant temperature effect. The diagonal lines is $y = x$.↩